

日本語音声におけるパワースペクトル因子の音声知覚上の役割

岸田, 拓也

<https://doi.org/10.15017/1931919>

出版情報 : 九州大学, 2017, 博士 (芸術工学), 課程博士
バージョン :
権利関係 :

氏 名 : 岸田 拓也

論 文 名 : 日本語音声におけるパワースペクトル因子の音声知覚上の役割

区 分 : 甲

論 文 内 容 の 要 旨

本研究では、連続的に発話された音声の臨界帯域幅で 20 帯域に分割し、各帯域のパワー変動を多変量データとする因子分析によって得られた因子がもつ、音声知覚上の役割を調べることを目的とした。

本研究では主成分分析とバリマックス回転を組み合わせる方法で因子分析を行った。因子分析によって得られた数個の因子から元の各帯域のパワー変動を再構成し、このパワー変動を実現する雑音駆動音声を合成することで、因子に含まれる情報だけで音声がどれだけ明瞭に聴きとれるのかを調べようとした。しかし、主成分分析はデータの重心を起点にして主成分を算出するため、得られた因子により元の各帯域のパワー変動を再構成しようとする、もともと無音を表すデータが無音のまま再構成されないという問題が起きる。この問題によって、ほとんどの場合、再合成された雑音駆動音声には定常的な雑音が生じる。そこで、本研究では主成分分析の起点を無音の点に移動させた、起点移動主成分分析を新しく提案することで上述の問題を解消した。分析 1 として、男性母語話者 5 名によって発話された日本語・イギリス英語・中国語(普通話)の 3 言語の音声について、臨界帯域ごとのパワー変動を計算し、これを起点移動主成分分析にかけて因子を得た。得られた因子は、通常の主成分分析から得られた因子と同等の構造を有し、かつ累積寄与率も両分析法で 2% 以下の差であった。

実験 1 では、起点移動主成分分析によって得られた因子の数を 1～9 まで増やしなが、因子から日本語の雑音駆動音声を再合成した。因子数をいくつまで増やせば日本語の雑音駆動音声が明瞭に聴きとれるのかを調べることを目的に、19～24 歳の 12 名(男性 6 名、女性 6 名)を実験参加者に、合成した雑音駆動音声の聴取実験を行った。実験参加者によって聴きとられた内容をモーラ単位で採点し、モーラ正答率を条件ごとに得たところ、再合成に用いる因子数が 2 個まででは実験参加者は音声の内容をほとんど聴きとることができなかったが、3 因子用いた音声では 70% 程度の正答率が得られ、4 因子以上を用いた音声では 80% を超える正答率が得られた。

ここで、起点移動主成分分析によって得られた因子から元の帯域のパワー変動を再構成しようとするとき、非負の値であるはずのパワー値が負の値として再構成されてしまうことがあり、このことが実験 1 の結果に影響を与えていた可能性が示唆された。そこで、起点移動主成分分析によって得られた因子の因子負荷量を、直交性を維持しながら非負値化するという方法で非負基底因子に変換した。分析 2 ではこの非負基底因子が変換前の因子に比べてどれだけ累積寄与率が下がるのかを調べた。その結果、因子数によって異なるが、最大 6% 程度累積寄与率が下がるということが分かった。

実験 2 では、非負基底因子を用いて実験 1 と同様の実験を、19～25 歳の 12 名(男性 6 名、女性 6 名)を実験参加者に行った。この実験によって、実験 1 での音声の再合成で負のパワー値が生じたことによる影響が大きいかどうかを確かめた。実験 1 と同じく、因子数が 2 から 3 に増えるときにモ

ーラ正答率が急上昇するという結果が得られ、実験1と実験2の結果に本質的な違いはないと分かった。また、4因子以上を用いたときに、モーラ正答率は90%を超えた。

実験1・2の結果から4因子までで十分に明瞭に音声を知覚するだけの情報が得られると分かったので、これらの4因子の間に音声知覚上の役割に違いがあるのかを調べるための実験を行った(実験3)。この実験では、4因子の内の1つの因子が与える各帯域のパワー変動の情報を定常的なパワーに置き換えて雑音駆動音声を再合成した。これによって1つの因子がもつ情報が除去された。除去された因子の違いにより、音声の明瞭度がどれだけ異なるのかを調べるために、20~24歳の10名(男性5名、女性5名)を実験参加者に聴取実験を行った。その結果、510~1480 Hzの周波数帯域に大きい因子負荷量もつ因子の情報が除去されたとき、正答率が36.8%まで下がることが分かった。それ以外の因子が除去された場合は、正答率は60.1~64.1%であった。このことから、4因子の中で最も重要な因子は、510~1480 Hzの周波数帯域に大きい因子負荷量をもつ因子であると分かった。

510~1480 Hzの周波数帯域に大きい因子負荷量をもつ因子は、3因子だけ取り出す因子分析で得た因子の中にも見つけることができた。そこで実験4では、2因子からなる因子、3因子からなる因子、4因子からなる因子のそれぞれの中で2因子以上を用いて(それ以外を除去して)雑音駆動音声を再合成し、聴取実験を行った。実験には19~23歳の16名(男性8名、女性8名)が参加した。その結果、因子数が同じ2個であっても、510~1480 Hzの周波数帯域に大きい因子負荷量をもつ因子が合成に用いられている条件の方が、正答率が高くなることが分かった。しかしながら因子数が2個では、高くても30%程度の正答率しか得られず、実験参加者は十分に明瞭に音声を知覚することができなかった。

よって、510~1480 Hzの周波数帯域のパワー変動の情報が含まれていて、かつ3帯域以上のパワー変動の情報が含まれていることが、雑音駆動音声を明瞭に知覚するための条件であるとした。また、この因子は、言語のリズムを知覚する手がかりを情報として含み、聴取者に音声を知覚する上での枠組みを与える役割をもつと結論づけた。

音声の周波数情報のもつ知覚の手がかりの冗長性に関して、合成音声を用いた聴取実験によって調べた研究と音響的特徴の統計的分析によって調べた研究とが結びつけられた。本研究で新たに得られた知見や手法を取り入れることで、音声知覚の仕組みの解明が進むと期待される。